Increasing Training Efficiency through Automated Learner Performance Evaluation

Robert Siegfried, Torsten Müller, Wilhelmine E. Dohmen
Aditerna GmbH
Otto-Hahn-Str. 13 B
85521 Riemerling
GERMANY

robert.siegfried@aditerna.de, torsten.mueller@aditerna.de, wilhelmine.dohmen@aditerna.de

ABSTRACT

We fight as we train. Providing warfighters with optimal training is therefore critical to succeeding within the 21st century military operating environment. However, current evaluation and analysis of training is not always optimal, mainly due to two factors: availability of highly trained subject matter experts as instructors and increasing data volumes generated from today's ever-evolving training environments that make comprehensive review and consistent trainee assessment cumbersome, if not impossible.

We present a two-pronged approach with our training information management system RAPTOR that automatically computes trainee performance metrics to relieve instructors from routine tasks and to enable instructors to provide more immediate and consistent assessments.

While the current system already helps instructors to focus more on advanced evaluation and in-depth feedback, future system expansion would help groups of trainees to self-evaluate and discuss training events with their peers. Quantitative analyses of movement, communication flows and share of conversation within a simulated scenario provide valuable hints and can also work as conversation starters in the group evaluation of a shared training event. This paper presents results from working with US and European forces, utilizing the approach in training of fighter pilots and military medical personnel.

1.0 CHALLENGES OF CURRENT TRAINING ASSESSMENT

In today's live and simulated training, Measures of Performance (MOPs) are used to analyze how well an individual or team performed, document the training, and track training progress and current readiness. Unfortunately, little analysis is done to determine if the chosen training scenario ensures the trainee masters the specific skills or if there is a more effective way to train the desired skills. Figure 1 conceptualizes today's typical process of training delivery and subsequent scenario improvement. [1]

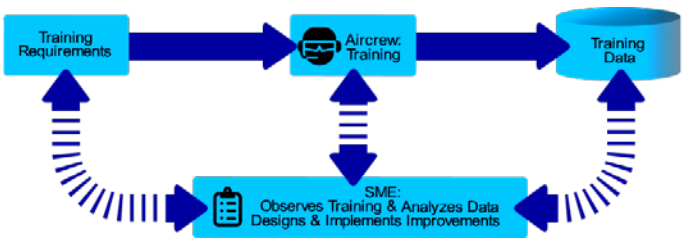

Figure 1. Typical training adaptation and improvement process today (simplified). [1]

Alas, the current training adaptation and improvement cycle faces multiple challenges. It is insufficient to deliver the amount and quality of training that is required to meet the demands in a rapidly changing security environment, including, but not limited to:

- Increasing data volumes raise magnitude and complexity of comprehensive reviews and consistent trainee assessments.
- Dependence on manual tasks (i.e., insufficient automation) and hence dependence on availability of suitably qualified and experienced personnel.
- Scarcity of said expert personnel.
- Subjective training assessment due to varying qualifications, experience, background etc. of instructors.
- Limited comparability of trainings due to varying quality, scope, documentation etc. of training assessments.
- Complexity of including information from additional sensors, e.g., to measure mental load, stress reactions etc. of trainees. [1]
- Stronger focus on soft skills within the military has been introducing more social elements to simulation trainings. [3] As these rely mostly on nuanced personal interactions and subjective judgments pure reliance on quantitative metrics and automated assessments yields little success.

The essential idea is to introduce automated training assessment into the training cycle that supports the instructors and trainees through objective trainee performance analysis and individualized training recommendations. Our key objective is to free up valuable time for instructors to ensure they can properly focus on those tasks that require expert judgment and cannot be automated.

We also want to point to recent research that shows better learning outcomes for students who were allowed to self-evaluate. [4] Empowerment and autonomy, even on a small scale, are also known to enhance sense of ownership, accountability, and pride in their work. Learning sciences connect these emotions to increased intrinsic motivation, creativity, satisfaction, and well-being.

A combination of automation and human-led evaluation that yields data-driven insights while enabling nuanced discussion and debriefing between trainees and instructors will achieve best results for training efficiency and knowledge retention.

2.0 AUTOMATED PERFORMANCE ASSESSMENT

Automated Performance Assessment (APA) refers to the use of advanced algorithms, analytics, and machine learning to objectively evaluate and measure the effectiveness of individual or group performance in training programs. This approach eliminates human bias and subjectivity and enables detailed feedback right after the training or even real-time, while the trainee is still in the simulation run, thereby optimizing training outcomes and resource allocation.

2.1 High Level Objectives

Automated Performance Assessment represents a paradigm shift in the way training effectiveness is measured and optimized. By leveraging the capabilities of advanced analytics, machine learning, and real-time data collection, automated assessment offers several advantages over traditional and/or manual evaluation methods. These systems eliminate the subjectivity associated with human evaluations, providing a more consistent and objective measure of trainee performance. Furthermore, they can process large volumes of data from complex training scenarios to deliver immediate and tailored feedback.

The main objectives of APA are:

- **Standardization** – One of the primary objectives is to standardize performance metrics across diverse training programs, allowing for more reliable comparisons and evaluations.
- **Support instructors** – Augmenting instructors in review sessions after a training run by providing objective, data-driven feedback on trainee's performance.
- **Support trainees** – Enabling trainees to access training results with or without the presence of an instructor.
- **Long-term tracking of performance** – Providing a long-term view on the past trainings of an individual to evaluate and visualize performance development over time.
- **Immediate Feedback** – Providing instant, real-time feedback to trainees is crucial for immediate correction and learning, facilitating a more dynamic and responsive training environment.
- **Data-Driven Decisions** – Empowering instructors and decision-makers with robust, data-driven insights into trainee performance and areas for improvement.
- **Resource Optimization** – By automating routine assessment tasks, these systems free up valuable human resources, allowing instructors and subject matter experts to focus on more nuanced and complex aspects of training and evaluation.
- **Scalability** – Scaling the assessment process effortlessly to accommodate small or large numbers of trainees, or to adapt to more complex or varied training scenarios, is a vital objective.
- **Continuous Improvement** – Leveraging collected data to iteratively refine and improve both the training programs and the assessment algorithms themselves, in a virtuous circle of continuous improvement.

2.2 Key Requirements

In the past years, while working with Automated Performance Assessment, we came across a variety of different requirements. The following subchapters highlight some of these requirements as they are generic and as such not only relevant for a limited audience and therefore are of specific interest for the M&S community as they provide valuable information to trainees and instructors.

2.2.1 Performance Metrics

Performance metrics within APA are quantifiable measures used to evaluate the proficiency and efficiency of a trainee in a given simulation. By analyzing data points such as response times, accuracy, decision-making pathways, and more, APA systems can provide an objective assessment of a trainee's capabilities. The data-driven nature of these metrics offers a level of consistency and objectivity seldom achievable by human evaluators.

However, different Communities of Interest (COIs) often have distinct performance metrics, shaped by unique guidelines, protocols, and manuals. For instance, the TCCC (Tactical Combat Casualty Care) requires different performance metrics in comparison to a flight manual from the Air Force.

Each of these documents has been crafted with deep expertise and understanding of the field they cater to. They are detailed and are often not compatible with metrics from another COI. This makes a "Universal APA" approach and system difficult to achieve. However, using modern approaches, e.g., a modular system design and technologies like Machine Learning (ML) and Large Learning Modules (LLM), helps bridging the gaps and make the next leap towards a COI-overarching APA. Evolving the APA alongside the ever-changing documents and metrics is key to keeping assessment results relevant.

While the promise of APA is immense, navigating the labyrinth of diverse performance metrics is a significant challenge. However, with the right strategies, it's possible to harness the full potential of APA across diverse training scenarios.

2.2.2 Training Recommendations

Analyzing past trainings not only provides insights on the trainee's past performance, but it can also be used for training recommendations for future trainings. Based on "currency" and "proficiency" requirements, these recommendations need to include both trainees' abilities and the organization's training standards like manuals and specifications.

While currency refers to the state of being up to date in a particular skill set, meaning that an individual or team has undergone the necessary training exercises or evaluations recently enough, proficiency refers to the measurable level of skill or competence a trainee has reached in a specific skill, task, or procedure.

This distinction is relevant also for APA as these are two key metrics to monitor and optimize training efforts. The two serve as a bridge between achieving training goals as a mere "done" checkmark (currency) to actionable insights in terms of skill level achieved (proficiency).

2.2.3 Video Replay

The APA analysis results are key for data-driven and objective debriefing. In addition to that, video replay augments debriefing by providing instructor and trainee with the possibility to replay specific scenes that were key moments in the simulation. These moments can be discussed and reviewed again as another object aspect in the debriefing. The video feed can be paused, rewound, played in slow-motion and forwarded from one key moment to another. This helps to dissect complex scenarios, examine decision-making processes, and identify both mistakes and successful strategies.

Seeing one's actions on screen can reinforce both good practices and reveal bad habits, helping trainees internalize the consequences of their actions and strategies. This visual reinforcement aids in the quicker mastery of complex skills.

When used in a group setting, video replays offer the chance for collective learning. Team members can point out details that others may have missed, fostering a collaborative learning environment and enhancing team cohesion.

Also, reviewing a video replay allows trainees and instructors to detach emotionally from the events, offering a calmer environment in which to analyze performance. This can be particularly beneficial after high-stakes or intense training scenarios.

By incorporating video replays into the debriefing process, the educational outcomes of simulation-based training runs can be significantly enhanced, making them a valuable tool for both trainees and instructors.

2.3 Market Overview

There is a variety of tools that have been developed to serve needs similar to those addressed by APA. These tools offer a range of functionalities, from real-time performance metrics to advanced analytics and simulation-based training assessments. Here's a brief overview of some of these tools and notable efforts:

- US Air Force (USAF) Secure Live Virtual Constructive Advanced Training Environment (SLATE)
- Generalized Intelligent Framework for Tutoring (GIFT) [5][2]

- Cervus XCaliber for marksmanship training [7]
- Aptima Learning Analytics
- Aditerna RAPTOR (see next section)

2.4 RAPTOR (in a Nutshell)

Thorough trainee evaluation requires subject matter experts (SMEs) to assess current performance and recommend future training that is optimized for each individual. Presently, with limited SME availability and large volumes of data generated from today's training environments, a comprehensive and consistent trainee assessment is impractical, if not impossible. Typically, this impacts the quality of analysis required to determine a trainee's actual proficiency, which in turn degrades the ability to select the most appropriate future training, which further degrades the ability to determine if a scenario can truly ensure that training objectives are mastered.

To overcome this problem, Aditerna has designed the system RAPTOR that ingests training data, calculates predetermined Measures of Performance and then provides scenario recommendations that ensure trainees master warfighting skills as intended (see Figure 2). By tracking trainees throughout their career, we can identify when a skill is mastered and how often trainees need to practice that skill to stay proficient.

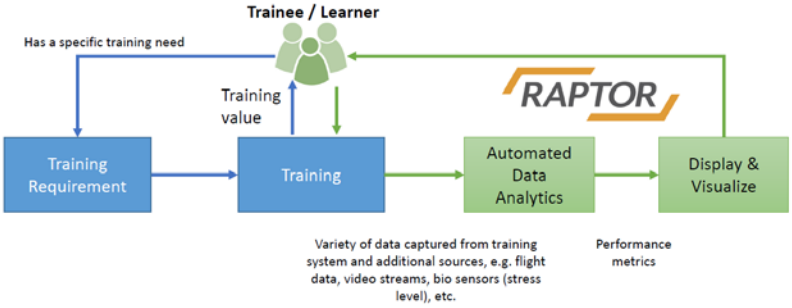

Figure 2. RAPTOR overview

RAPTOR is currently being developed further as part of the "Virtual Reality Trauma Simulator" (VireTS) project within the European Defence Industrial Development Programme (EDIDP) as a solution to analyze data from a medical Virtual Reality (VR) simulation. [7] While the initial focus is on Tactical Combat Casualty Care (TCCC) trainings of military medics and military first responders, the scope of the project should widen to eventually include comprehensive military medical care.

RAPTOR transforms training data from VR applications and simulators into actionable intelligence by offering consistent trainee performance measurement and long-term training data analysis. Providing automated feedback and recommendations for future training increases both training effectiveness and efficiency.

3.0 EXPANSION AND FUTURE OPPORTUNITIES

3.1 Future Opportunities for Individual Learner Evaluation

Virtual Reality (VR) has revolutionized the way we approach training and education, offering immersive and interactive experiences that can simulate real-world scenarios. It contributes to training effectiveness by providing objective metrics, opportunities for practice, immediate feedback, and individualized learning

paths. The impact on training evaluation cannot be overstated as it brings together high immersion, controllable and repeatable training conditions and as vast treasure of data including trainee decisions and movements as well as response times, accuracy, and completion rate.

3.1.1 Real-Time Assessment and Adaptive Training

Real-time feedback in training offers trainees an opportunity to adjust their approach, correct mistakes and reinforce positive behaviors. Combined with a dynamic training management system and adaptable scenarios real-time assessment also allows for an adjustable training environment that can offer additional guidance or provide extra practice opportunities and might even influence the difficulty level of training scenarios.

Real-time assessment is further enhanced by the integration of biosensors to monitor physical data and include an objective stress level assessment which can be used to ensure trainee safety and health. Furthermore, it enables customized training interventions and even biofeedback training for trainees to control stress levels, improve focus, and optimize performance.

3.1.2 Integration of Biosensors

Significant advancements in technology, improved sensor capabilities, and increased accessibility to these technologies have made the integration of biosensors into training much easier. Even though higher costs combined with the complexities of hardware and software integration have somewhat hindered large-scale introduction of biosensors in training we expect comprehensive implementation in the near future.

Integrating biosensors into training programs maximizes the benefits of training while also contributing to better preparation for real-world challenges and missions. Providing trainees with personalized, data-driven guidance from biosensors optimizes training efficiency and effectiveness as it helps trainees improve their performance, manage stress, and overall enhance their own training experience.

3.1.3 Spatial Analysis

Spatial positioning during simulation training can offer valuable insights into trainee’s knowledge and confidence, e.g., walking fast and direct to medicine stock as opposed to wandering slowly and haltingly searching for the right resources. The data received can also be used to optimize path planning both in VR (for future simulations) and in reality, to find the most suitable path to any given goal and to research the best location for tools to be used during training or in real life.

3.2 Additional Offers for Team Performance Evaluation and Group Assessment

Some of the most crucial skills for warfighters are group-oriented and emphasize teamwork, clear communication, leadership, and situational awareness. [9][10] As mentioned before it is extremely challenging to automate performance assessement for social skills like these. Nevertheless we aim to provide instructors and trainees with hints and thought-provoking impulses based on trainees‘ simulation experience and automatically calculated by our system to support the evaluation and development of social skills.

Using motion profiles, communication patterns, and voice analysis, trainers and educators can gain valuable insights into group dynamics, individual contributions, and areas for improvement. This in-depth analysis not only enhances the effectiveness of VR training but also enables a more nuanced approach to developing interpersonal skills and teamwork. Analyzing group interactions training provides a multi-dimensional understanding of participant behavior as it helps reinforce these skills.

3.2.1 Motion Profiles

Motion profiles refer to trainees' movements and body language in a VR environment. Data includes gestures, postures, and spatial positioning. They can offer valuable insights into group dynamics and individual behavior during training sessions such as patterns of leadership, submissiveness, or collaboration. Spatial analysis can also help identify the optimal positioning of team members during critical procedures, ensuring efficient communication and coordination.

3.2.2 Communication Analysis

Trainers can define metrics for communication, such as the frequency and quality of verbal interactions, the use of structured communication protocols (if applicable), and the accuracy of conveyed information. Evaluations can focus on whether trainees become more effective in conveying critical information, listening actively to their team members, and maintaining clear and concise communication during high-stress scenarios. There is also research into evaluating trainees' situational awareness through communication analysis promising even more insights into participants' "inner workings". [11]

3.2.3 Share of Conversation Time

Voice recognition technology can detect patterns such as who speaks the most, who listens actively, and the tone of conversations. For instance, a participant who speaks most while consistently interrupting others might be exhibiting dominant behavior.

3.2.4 Automated Support Functions for After Action Reports and Debriefing

A holistic approach of including automated feedback and human interaction in debriefing with trainers is considered optimal for learning evaluation and training efficiency. But, as we experience an increasing scarcity of SMEs and other educators, we want to explore how a sophisticated Training Information Management System (TIMS) can support the facilitation of After-Action Reviews (AARs) for a group of trainees even in the absence of instructors. However, to ensure a well-rounded training assessment trainees should ideally receive guidance on how to use the TIMS effectively for AAR purposes and additionally self-guided AARs should be periodically complemented with instructor led AARs.

Partly automating After-Action Reports (AARs) can streamline the process, reduce manual effort, and ensure consistency in reporting. More importantly it supports organizations in fostering a culture of continuous improvement by empowering team members to assess and enhance group dynamics by themselves. This approach not only helps trainees develop essential skills but also ensures that these principles are consistently applied and refined in high-stress environments.

The main areas where TIMS can support self-guided AARs to empower trainees and help overcome shortages of instructors:

- **Data Access and Visualization –** TIMS provides trainees with access to the collected training data, including performance metrics, communication logs, and simulation outcomes. TIMS can automatically collate data to generate charts, graphs, and to present information in a visually appealing and comprehensible manner. TIMS' visualization tools enable trainees to explore data visually, helping them to identify trends, patterns and anomalies independently.
- **Performance Metrics** - Trainees can assess their individual and team performance against predefined KPIs. Access to performance data allows trainees to identify strengths and weaknesses as TIMS presents performance metrics calculated from the training data.
- **Customized AAR Templates** - TIMS can include predefined and customized AAR report templates to use as a starting point for self-guided AARs. Trainees can customize these templates to align with their

specific training objectives and areas of interest.

- **Version Control and Historical Data** - Enabling trainees to track changes and improvements in their AARs over time is done by TIMS maintaining version control for AARs. Historical AAR data serves as a reference for tracking progress and assessing the evolution of training outcomes.
- **Event Logs and Timeline** - Event logs and timelines recorded by TIMS help trainees review and reconstruct the sequence of actions and decisions made during the training. This feature helps trainees gain a comprehensive understanding of the training scenario.
- **Participant Feedback Integration** - TIMS can be adapted to include survey and feedback tools so that trainees can review feedback from themselves and their peers. This feedback provides valuable insights into the training experience and areas for improvement both for the training groups as well as individual trainees.
- **Collaboration and Discussion** - Trainees can use collaboration features within TIMS to engage in group discussions and share insights with their peers. Group discussions support trainees to collectively identify lessons learned and generate recommendations.
- **Actionable Insights and Recommendations** - Automated analysis based on available training data can generate actionable insights and recommendations. Trainees can review and adapt these to guide their self-guided AARs.

While we hope to make evident the value of automation in support of AARs and generally to enhance training effectiveness the human element remains crucial, especially in the interpretation of results, the identification of lessons learned, and the formulation of actionable recommendations. Automation complements human expertise and can expedite the reporting process, but the insights and contextual understanding provided by experienced personnel remain invaluable in AARs whether facilitated by an instructor or group-led. [12]

4.0 CONCLUSION AND OUTLOOK

4.1 Conclusion – Automated Performance Assessment (APA) to the Front

In conclusion, the pursuit of increasing training efficiency through automated performance assessment (APA) represents a pivotal advancement in educating and training professional warfighters.

This research has illuminated the potential of APA to enhance not only the effectiveness of training programs but also the overall learning experience. The ability to provide timely, personalized feedback and adapt instructional content in real-time offers a promising avenue for improving learner outcomes and engagement.

To overcome critical challenges of current training assessment (such as the raising complexity and magnitude of comprehensive reviews due to increasing data volumes, dependence on manual tasks, limited availability of instructors, subjectivity of assessments, complexity of including additional information from new sensors), an increased use of automated performance assessment seems to be a promising solution approach. While freeing up valuable instructor time, automated performance assessment helps to generate objective (and comparable) assessments for each trainee and enables the integration of a vast array of data sources.

Various automated performance assessment tools are currently being developed and have reached different maturity levels. Although the general benefit of such tools has been demonstrated in many different applications, they are still far from being routinely used. The utilization of APA can and should be increased by investing time and effort into cutting-edge technologies and data-driven approaches to further enhance

training efficiency and effectiveness.

However, it is crucial to acknowledge that this pursuit is not without its challenges, including issues related to the ongoing refinement of automated evaluation systems and smoother integration of external data collectors (like biosensors). Therefore, as we move forward in implementing these technologies, a commitment to continuous improvement, and the integration of human expertise remains paramount. In summary, the journey towards increasing training efficiency through automated learner performance evaluation is an exciting one, marked by transformative potential. By harnessing the power of automation and leveraging data-driven insights, we can pave the way for more effective and personalized learning experiences, ultimately advancing our collective pursuit of knowledge and skill development in modern warfare.

4.2 Outlook – Steps Toward Automated Support for Debriefing and After-Action Report

While we have made significant strides in harnessing the potential of automated learner performance evaluation, the path forward is marked by several exciting opportunities and challenges. To fully realize the vision of enhanced training efficiency, we must consider the following key aspects:

- **Supporting Instructors through Automation** - Instructors play a pivotal role in the learning process. Future developments should prioritize the creation of automated tools that assist educators in interpreting learner performance data effectively. This support can aid instructors in tailoring their teaching methods and interventions, ultimately fostering a more engaging and responsive learning environment.

- **Empowering Trainee Groups for Independent AARs** - The concept of debriefing and AARs as means of reflection and learning is invaluable. Going forward, there is a pressing need to empower trainee groups to conduct AARs autonomously when circumstances warrant it. Automated systems supported by ML and LLM can provide the necessary data and guidance for these reviews, encouraging self-directed learning while enhancing collaboration among learners.

- **Realizing Continuous Improvement** - The journey toward training efficiency is an iterative process. To stay on the cutting edge, we must embrace a culture of continuous improvement. This involves refining automated evaluation systems based on user feedback, emerging best practices, and advances in technology. By doing so, we can ensure that these tools remain relevant and effective.

- **Scaling Personalization** - Personalization of training experiences offers tremendous benefits. The challenge lies in scaling these approaches to accommodate a diverse range of learners, including those in large online courses. Further exploration of adaptive algorithms and scalable personalization techniques is needed.

- **Cross-Sector Collaboration** - Collaboration across sectors, including academia, industry, and government, is essential. This collective effort can drive innovation, share insights, and facilitate the development of standards and best practices for automated learner performance evaluation.

- **Ethical and Inclusive Automation** - As automation becomes more integrated into training and evaluation, addressing ethical concerns and promoting inclusivity is paramount. Future work should prioritize research on mitigating algorithmic bias, protecting learner data privacy, and ensuring that automated systems do not inadvertently disadvantage certain groups of learners.

In summary, the outlook for increasing training efficiency through automated performance assessment enriched with additional information on social cues is a compelling one. It encompasses not only the

optimization of technology but also the empowerment of instructors and learners alike while at the same time acknowledging that social competence may determine whether a particular mission succeeds or fails. [13]

REFERENCES

[1] Robert Siegfried, Torsten Müller, Mark Burgess, Krzysztof Rechowicz: "Providing Better Feedback to Aviators through Automated Human Performance Analysis", I/ITSEC 2021, Orlando, USA. Paper No. 21238.

[2] Christopher D. Wickens: "Mental Workload: Assessment, Prediction and Consequences" in: "Human Mental Workload: Models and Applications. First International Symposium, H-WORKLOAD 2017", Pages 27-57

[3] Tom Bijlsma, Carla Broer (2016): "Crew Resource Management as Shield and Spear for Safety and Security" in: Robert Beeres, Gwendolyn Bakx, Erik de Waard & Sebastiaan Rietjens (2016): "NL ARMS Netherlands Annual Review of Military Studies"

[4] Patrizia Kündig, Franziska Tschan, Norbert K. Semmer, Camille Morgenthaler, Jasmin Zimmermann, Eliane Holzer, Simon Andreas Huber, Sabine Hunziker & Marsch, S. (2020): „More than experience: a post-task reflection intervention among team members enhances performance in student teams confronted with a simulated resuscitation task-a prospective randomised trial." In: „BMJ simulation & technology enhanced learning" Vol. 6(2), pp. 81–86.

[5] Robert Sottilare, Ross Hoehn, Alyssa Tanaka: "Adapting existing simulation architectures to enhance tailored instruction", I/ITSEC 2019, Orlando, USA. Paper No. 19239.

[6] Randall Spain, Jonathan Rowe, Benjamin Goldberg, Bob Pokorny, James Lester: "Enhancing Learning Outcomes Through Adaptive Remediation with GIFT", I/ITSEC 2019, Orlando, USA. Paper 19275.

[7] https://cervus.ai

[8] https://defence-industry-space.ec.europa.eu/system/files/2021-06/EDIDP2020_factsheet_SME_VIRETS.pdf

[9] Hilde van Ginkel: "Crew Management Training and Research in Military Organizations" in "Advances in Military Navigational Technology" Vol. 2010, pp. 221-240.

[10] Paul O'Connor, Robert G. Hahn, Robert Nullmeyer (2010): "The Military Perspective" in Barbara G. Kanki, Robert L. Helmreich, José Anca (2010): "Crew Resource Management", pp. 445-466.

[11] Peter W. Foltz, Cheryl Bolstad, Haydee M. Cuevas, Marita Franzke, Mark Rosenstein & Anthony Costello (2008): "Measuring Situation Awareness through Automated Communication Analysis" in Michael P. Letsky, Norman W. Warner, Stephen M. Fiore & C.A.P. Smith (2008): "Macrocognition in Teams: Theories and Methodologies", pp. 259-275.

[12] Sylvain Boet, M Dylan Bould, Bharat Sharma, Scott Revees, Viren N Naik, Emmanuel Triby &Teodor Grantcharov (2013): "Within-team debriefing versus instructor-led debriefing for simulation-based education: a randomized controlled trial" in: "Annals of surgery" Vol. *258*(1), pp. 53–58.

[13] ARMY LEADERSHIP AND THE PROFESSION: Army Doctrine Publication 6-22, downloaded Sept.22nd 2023 from https://irp.fas.org/doddir/army/adp6_22.pdf